DOCUMENT RESUME

ED 359 935 IR 016 155

AUTHOR Richards, William R.
TITLE An Application of Digitized Speech in Hypermedia.
PUB DATE 93
NOTE 22p.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Audio Equipment; Audiovisual Communications; *College Students; Computer Assisted Instruction; Computer Simulation; Computer Software; *Computer Software Development; Formative Evaluation; Higher Education; *Hypermedia; *Interactive Video; *Speech
IDENTIFIERS Digital Data

ABSTRACT

Hypermedia applications have presented information through a variety of visual media, but the aural channel for information delivery has not been well developed. To reduce the likelihood of overloading the visual channel of communication in a program that presents a great deal of information through graphic illustration and animation, the hypermedia program "Field Kit Workshop" (FKW) uses speech as the primary means of delivering verbal information. FKW is an interactive simulation that introduces students to operating features of professional video production equipment. A formative evaluation was conducted with 13 volunteer students of video or audio production to explore user response to speech as used in FKW, and to help guide implementation of speech in the program's final design. Results suggest that speech was accepted by users within a program that is well-designed overall, and in which the design takes into account the special strengths and weaknesses of speech as a medium for delivery. Sixteen figures illustrate the discussion, and an appendix presents an excerpt of a program script for FKW. (Contains 8 references.) (SLD)

AN APPLICATION OF DIGITIZED SPEECH IN HYPERMEDIA

William R. Richards

ABSTRACT

Today's technology has made digital sampling of audio for computer storage and playback a "desktop" venture. But the widely available capability has not resulted in widespread application. Perhaps a first step in finding a productive use for audio in hypermedia is to reduce our dependence on text displays as the accepted mode for presenting verbal information.

To reduce the likelihood of overloading the visual channel of communication in a program that presents a great deal of information through graphic illustration and animation, the hypermedia program, "Field Kit Workshop" (FKW), uses speech as the primary means of delivering verbal information. FKW is an interactive simulation that introduces students to operating features of professional video production equipment.

Formative evaluation was conducted to explore user response to speech as it was used in FKW, and to help guide the implementation of speech within the program's final design. This study found that speech was accepted by users within a program that is well-designed overall, and in which the design takes into account the special strengths and weaknesses of speech as a medium for delivery.

INTRODUCTION

Less than a decade ago, computer-based instruction was almost exclusively presented through on-screen text. From beginnings in this text-only environment, computer-based instruction has evolved into today's hypermedia. In practice, hypermedia applications have presented information through a variety of visual media, but the aural channel for information delivery has not been well developed. Locatis, *et al*, writing as recently as 1990, define hypermedia as composed of three subsets: hypertext, hypergraphics, and hypervideo (Locatis, 1990). This definition describes visual media -- no mention is made of "hypersound."

Today's computer technology has made digital sampling of audio for computer storage and playback a "desktop" venture. But the widely available capability has not resulted in widespread application. As one columnist writes in the computer press, "nobody's even figured out how to use sound productively, and it's been built into the Mac for over a year now" (Zilber, 1992). Perhaps a first step in finding a productive use for audio in hypermedia is to reduce our dependence on text displays as the accepted medium for presenting verbal information.

For this project, a hypermedia program was created which uses speech as the primary means of delivering verbal information. Designed as an introductory step in training students to operate a professional-grade portable video tape recorder, "Field Kit Workshop" is a program that uses speech within a visual context of detailed images, both still and animated, and a rich audio context of realistic sound effects and music. Formative evaluation was conducted to explore user response to speech as it was used in "Field Kit Workshop," and to help guide the implementation of speech within the final design of the program.

LITERATURE

Information can be presented to the user of hypermedia through a variety of visual and auditory means. The most common mode of presentation in computer-based instruction has been text displays, with graphics being the next most common. Sound as a presentation mode is an option infrequently used. When sound has been used, the sounds have often been nothing more than "primitive sound effects, such as beeps or explosions" (Alessi & Trollip, 1991).

The chief motivation for delivering verbal information through speech rather than text in the current project is to reduce the likelihood of overloading the visual channel of communication in a program that presents a great deal of information through graphic illustration and animation. Fleming and Levie's analysis of studies from a wide range of disciplines supports the notion that speech can be more effective than text in such situations:

> "Capacity [to perceive] appears to be larger where two modalities are utilized (audition and vision) rather than one. Two tasks involving the visual modality, for instance, will interfere more than where one involves the visual and one the auditory modality" (Fleming & Levie, 1978).

This makes sense when one considers that it is much easier to look at an illustration while listening to narration than it is to look at an illustration while reading text. Fleming and Levie caution that discrepancies across two modes can impede learning, and that "excessive redundancy" across two modes of delivery, such as text and speech that deliver identical words, "may induce boredom or inattention to one modality" (Fleming & Levie, 1978).

Fleming and Levie point out that receiving information through speech can put great demands on short term memory -- since the meaning of a sentence may not be apparent until it is completely delivered -- and offer the recommendation that spoken phrases be kept short. Fleming and Levie also state that conversational speech (as opposed to written text that is read aloud) seems naturally divided into phrases that present no difficulty in perception (Fleming & Levie, 1978).

Although the need to present information in small units may seem to limit the usefulness of speech in computer-based instruction, it does not automatically follow that text is a superior mode of presentation; a consensus among hypermedia designers is that on-screen text also should be presented in small information units, commonly called "chunks" (Carlson, 1990; Failo & DeBloois, 1988; Knuth & Brush, 1990). It may be that the nature of on-screen presentation puts text on nearly even footing with speech regarding the amount of information that can best be presented per unit.

Rate of speech in words per minute (wpm) is a characteristic of narration that can affect intelligibility. Marics and Williges refer to studies that examined rates of speech, in which conversational speech is typically found to be at a rate of around 180 wpm, with compressed natural speech being understandable at 280 wpm (Marics & Williges, 1988).

Marics and Williges also found that subjects transcribing from speech recalled words from the ends of messages more accurately than from the beginning of messages, and that errors in receiving information through speech can be reduced if the user has the option of repeating the message (Marics & Williges, 1988).

RESEARCH AND DESIGN QUESTIONS

The current study came about as the result of design challenges that were raised during early development of "Field Kit Workshop" (FKW), an interactive program intended to provide an introduction to the operation of video production equipment. The program design relied heavily on detailed visual images -- images that quickly became cluttered in early versions as text

overlays were added to guide the user through the program and provide information about operating controls. A possible solution to "visual overload" presented itself. Perhaps speech, rather than text, could be used to guide the student through the steps of operating the equipment.

Review of the literature supported the notion that speech might be used effectively in some hypermedia programming, and the decision was made to incorporate speech into the design of the proposed program. It was also decided to conduct formative evaluation to help determine whether speech display was appropriate for "Field Kit Workshop," and to guide the way in which speech display would be applied in the final version of the program.

One question to be resolved was whether speech would be effective in providing the brief tutorial and procedural information that comprised the verbal component of FKW. Doubts that the literature raises about the listener's ability to retain spoken information make this question an important one in deciding to use speech.

The literature cited above points out the need for speech displays to be repeatable by the user, as an aid to understanding. What is an effective design for repeating speech that can compensate for the shortcomings speech might have in terms of intelligibility and retention?

Another key question relates to user acceptance. Given that verbal information has traditionally been delivered as text in hypermedia and other forms of computer-based instruction, will users be open to receiving information in the form of computer-delivered speech?

METHOD

PRODUCTION DESIGN

An instructional hypermedia program was produced that uses digitized speech to present informational content. The subject of the program is the operation of a professional-grade videotape recorder for use in field production. The program, "Field Kit Workshop," was designed for presentation on the Apple Macintosh II family of computers, using the software program, HyperCard.

Instructional Goals

The instructional goal of the program, "Field Kit Workshop," is to familiarize the student with the basic operating features of the Sony BVU-150 video tape recorder (VTR), in preparation for a controlled, hands-on exercise that involves setting up a field production kit for an interview. Hypermedia presentation was seen as a way to provide more detailed information about the equipment than was feasible in the lecture/demonstration format of the typical equipment introduction; at the same time, the interactive, "hands-on" feel of hypermedia would make this detailed information more meaningful to the student.

Program Content

The program introduces the student to the Sony BVU-150 video tape recorder by guiding the student through the procedural steps necessary to prepare the VTR for recording an interview.

The body of the program can be divided into nine segments that cover the operating functions of the Sony BVU-150 video tape recorder. Select segments are described in Figures 1 through 5. Included in the program is a series of introductory modules that describe program operation and navigation. No data relating to user behavior is recorded during these introductory modules.

Figure 1. Introduction to the BVU-150
In this very brief introduction to the BVU-150 videotape recorder (VTR), the workshop instructor describes the unit in terms of advanced features, such as high resolution recording and a built-in time code generator.

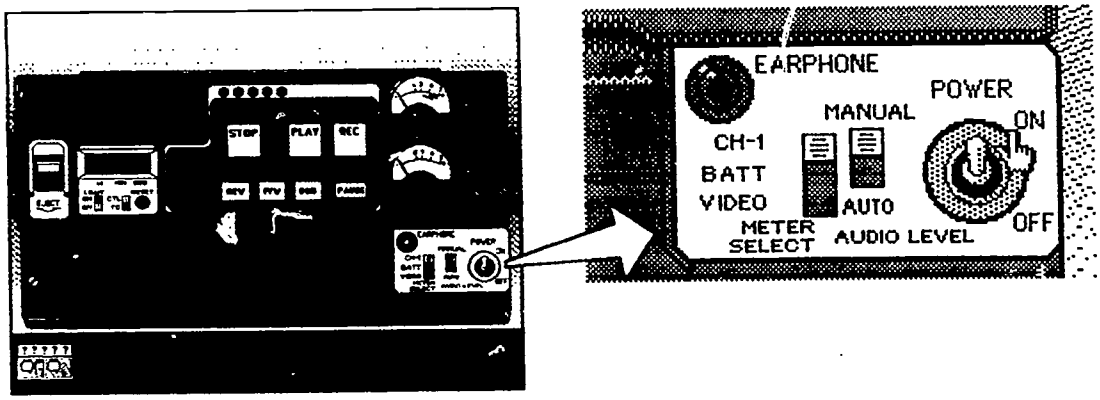

Figure 2. Powering Up
Here the trainee is directed to turn the deck power on. The trainee learns that the tape counter serves as a power-on indicator, and is then led through the steps of checking the charge on the battery using the VU meter for audio channel one.

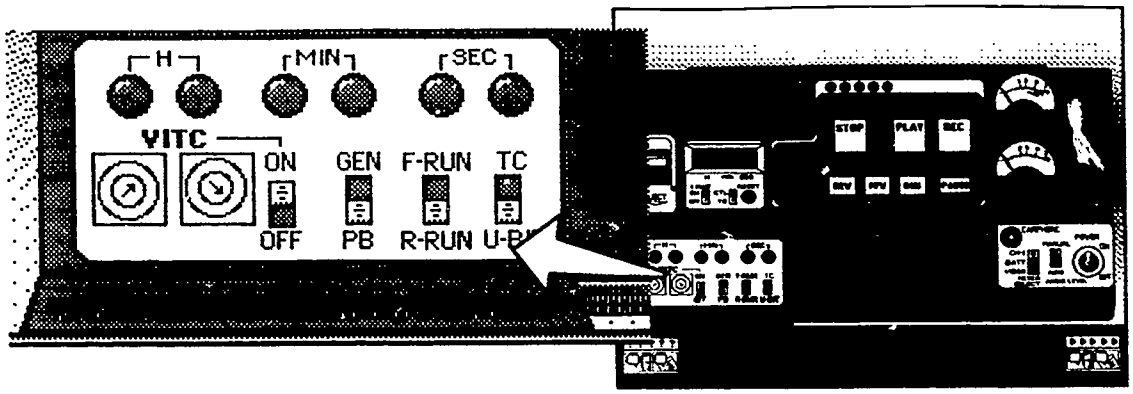

Figure 3. The Time Code Panel.
The trainee switches the tape counter into time code display, and the instructor introduces the control panel used for setting the time code generator. The instructor gives a very brief explanation of four switches that set parameters for recording time code; the trainee sets these switches, and sets the starting hours, minutes and seconds for the time code.

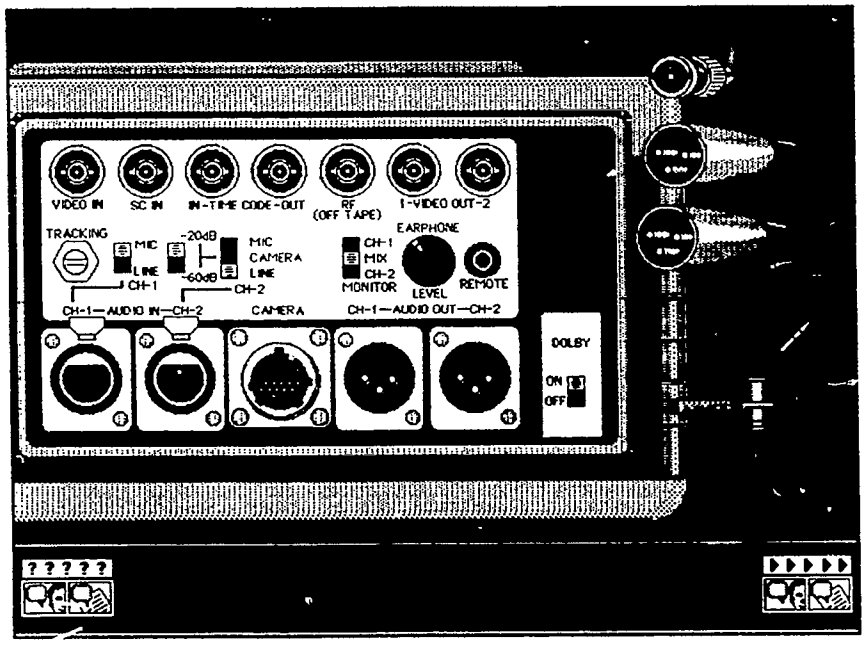

Figure 4 Connecting Cables.
The trainee is directed to the VTR connector panel, located on the side of the deck opposite the battery compartment. Here the instructor leads the trainee through the necessary cable connections: a lavalier (or "tie-tac") microphone is connected to an audio cable, and then to audio channel two; the output of the time code generator is patched into audio channel one with an adapter cable; and the camera cable is connected. Proper line/mic input levels are set with the appropriate switches, and the switch for Dolby noise reduction is turned off.

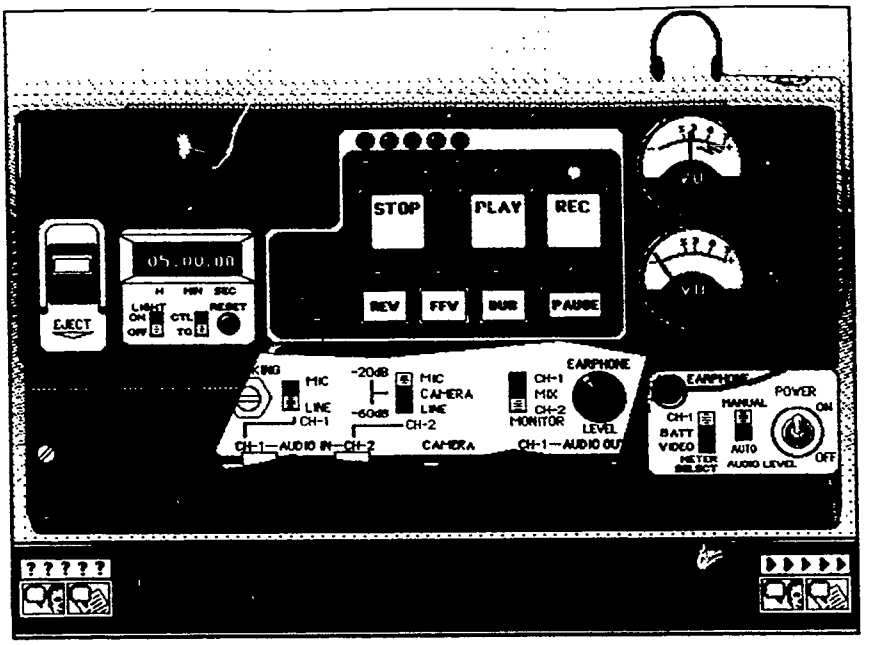

Figure 5. Setting Audio Levels.
The trainee returns to the VTR control panel, and adjusts the audio level of the time code signal in channel one; checks the audio level for the mic in channel two; and uses the VU meter for channel one to check the video signal from the camera. Here the trainee also learns to adjust the gain and the output (ch1,ch2 or MIX) for earphone monitoring.

Program Structure

The basic structure of FKW is linear, since the student is guided on a fixed path through a standard procedure made up of a series of specific steps. Within this linear structure, information had to be structured in such a way that would provide for the greatest understandbility of the information as spoken, and that would make it possible to offer the user options to repeat spoken information as necessary.

In keeping with the vocabulary of hypermedia, each unit of information within the program will be referred to as a *node*. In FKW, a node of information is typically composed of several smaller parts: one or more sentences of verbal information relating to a single fact; a static or animated visual which illustrates or complements that verbal information; and a specific program response to user manipulation of virtual controls. From the user's standpoint, a node consists of everything that lies between two navigation decisions.

The prototype version of the program contains forty-seven nodes of information. Thirty-one of the forty-seven nodes require the user to perform some specific action as a part of the procedure for preparing the VTR to record. Within one of these *action nodes*, the user is directed to perform some action on-screen. When the correct action is performed, additional information may be presented, or the node may be complete.

Figure 6 depicts an action node in its most basic form. When the user sends a navigation command to CONTINUE, the node begins with a sentence display that provides tutorial information -- in this case, the proper setting for the audio level in channel one. This tutorial information is immediately followed by a procedural instruction -- a sentence that directs the user to turn a certain dial on the control panel. At this point the user is given the option to REPEAT the procedural instruction, if necessary. The user then performs the action as instructed. The result of the user's action in this node is a new setting on the simulated VU meter. With the

correct setting, the node is complete, and the user has reached another navigation point. Here the user can choose to REPEAT this node or to CONTINUE to the next.

visual	audio
navigate	
VU meter reads at maximum	"The timecode signal is way too hot. It should be between -5 db and -3 db."
	"Adjust the level for audio channel One to put the timecode signal midway between -5 and -3."
action: dial CH 1 counterclockwise	
needle adjusts to -4	sfx: dial
navigate	
	"While you're still at the meter for channel One, check to see that the deck is getting a good VIDEO signal from the camera.

Node Begins
Tutorial Information
Procedural Instruction
User Action & Result
Node Ends

Figure 6. An Action Node.

Presentation Mode

In the design of the prototype version of the program, the user makes a choice of Speech Only presentation or Speech & Text presentation each time a navigation decision is made. This means that the user is choosing from one of four options: (1) REPEAT, Speech Only; (2) REPEAT, Speech & Text; (3) CONTINUE, Speech Only; or (4) CONTINUE, Speech & Text. Figure 7 illustrates the control panels that offer the user these four choices. Each of the two control panels on the bottom of the screen has icons representing the Speech Only option, and the Speech & Text option.

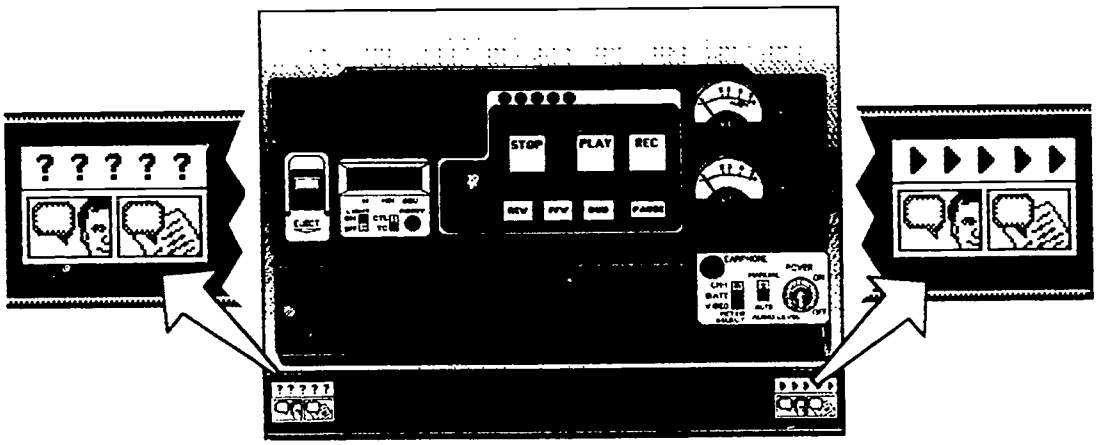

Figure 7. Navigation Panels in the Control Bar.

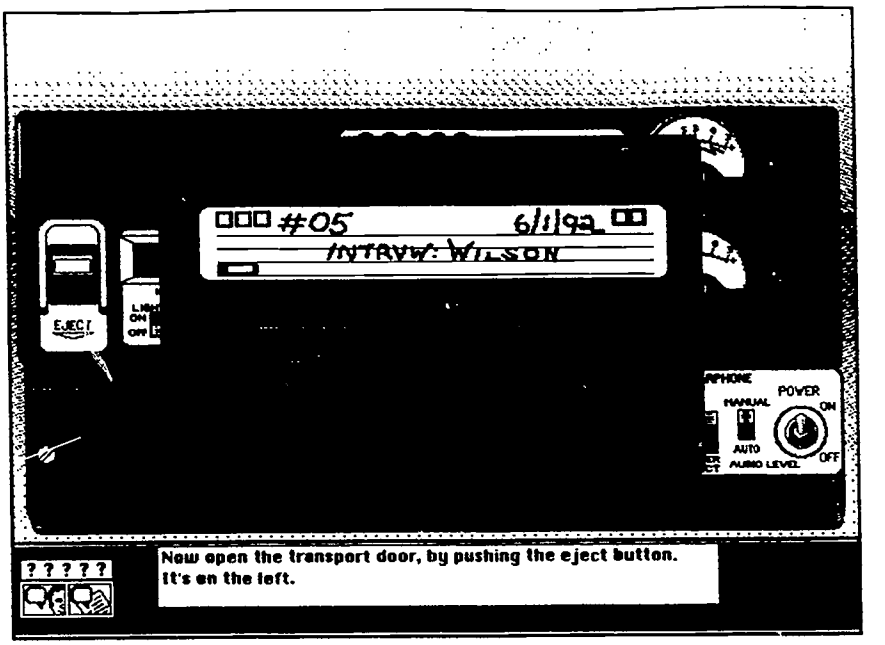

Figure 8. The Text Window in the Control Bar

When the Speech & Text option for presentation is selected, a "Text Window" appears in the center of the control bar. The text window contains the exact text as spoken by the narrator (see Figure 8).

Speech

Applying speech effectively in the program, "Field Kit Workshop," meant considering a range of characteristics of delivery, including scripting, recording quality, and rate of speech.

The program script for the "instructor" had to be written to be spoken rather than read. Syntax and diction were crafted to achieve a conversational tone. This generally meant breaking long sentences into shorter ones, using connecting words, and avoiding formal-sounding words and phrases. The program "instructor" uses the pronouns *you* and *I* to maintain the natural, conversational feel of the program.

The instructor's narration was recorded using a studio-grade microphone, a Sennheiser MD 421 U-5. This microphone was selected for its ability to capture lower frequencies that lend warmth to the recorded voice. All voice recordings were sampled at a rate of 11 kHz. A higher sampling rate of 22 kHz would have been preferred, but there was simply not enough disc storage space available. As it was, slightly over twelve minutes of voice recordings for the program required 8.4 MB of storage.

For FKW, it was decided that 200 wpm would be the target rate of speech for the narrator; close to the conversational rate of speech of 180 wpm, to maintain the conversational feel, but a little faster for the sake of keeping the program pace up. The actual average rate of speech in the program is 205 wpm.

Speech as Negative Feedback

At any given action point in the program, there is only one correct response that the trainee can make. When the user makes an incorrect response -- flipping the wrong switch, or connecting a cable to the wrong place -- FKW provides two types of "negative feedback": one, the attempted

action can't be completed (the switch doesn't respond, for example); and two, the program uses speech to tell the trainee that the action is incorrect.

Each time the user attempts an incorrect action in FKW, the instructor's voice delivers one of four messages, selected at random: "No," "Sorry," "Try Again," or "Sorry, Try Again." The variety of responses and the random element help to maintain the conversational feel of the program.

The Audio Environment

In addition to speech, the audio environment was inhabited by sound effects and music. Because these other sounds serve functions within the program which are intended to support information presented through speech, it is important to provide some description of these other audio elements.

Sound Effects

More than twenty sounds produced by the Sony VTR in operation were recorded to be used as sound effects within the program. The click of a switch, the spring of the tape eject mechanism, the distinctive sound of the tape being threaded around the tape head -- these and other equipment sounds were recorded at the maximum sampling rate of 22 kHz to maintain a high rate of realism. Slightly over one minute of VTR sounds occupy almost 2 MB of disc storage space. The sounds provide a natural way to give users audio feedback as they click switches, etc.; and lend realism to the program to enhance transfer of learning.

Music

The theme and incidental music which appears throughout FKW is provided by a single instrument, an acoustic bass, played in an improvisational jazz style. Additional music is provided by a basic drum set made up of kick drum, snare, tom-toms, hi-hat and cymbals. These sampled sounds are played back as themes and cues according to routines scripted in HyperCard's authoring language, HyperTalk.

The acoustic bass theme and incidental music accompany scene transitions within the program, and are used to "bracket" narration in introductory and review segments of the program. Occasionally a short phrase is used in conjunction with an animated, on-screen "pointer" to help draw attention to some visual detail in illustration or animation. Any of a variety of drumbeats announce the appearance of the CONTINUE navigation panel, and with it the need for the user to make a navigation decision to either continue or repeat.

EVALUATION DESIGN

Evaluation of the program was designed to explore how students use and respond to digitized speech as a mode of presentation in hypermedia. One aim of the evaluation was to gauge user response to and acceptance of speech as a means of delivery in the FKW program. A second aim was to gather information about decisions users make when given a choice between presentation modes. This information would be used to plan the design of a complete and final version of the "Field Kit Workshop" simulation.

Evaluation of "Field Kit Workshop" was essentially formative, intended to determine if delivery of verbal content by speech was appropriate to the specific needs of this program in terms of effectiveness and user acceptance. Questions explored included: Do students take advantage of the option to repeat speech displays? Do students desire on-screen text displays as a complement to speech displays? Can it be demonstrated that a program such as "Field Kit Workshop" can be designed to effectively deliver verbal information through the medium of speech?

Sample

The program was tested with a non-probability sample comprised of students who responded to posted notices and in-class requests for study participants. All participants were either currently enrolled in or had completed basic video or audio production coursework. A total of thirteen volunteer subjects took part in the study. The small sample size was appropriate to the nature of the study as formative evaluation.

Instruments

One instrument of measurement was a record of presentation choices made within the program by each student. Each user command to CONTINUE or REPEAT was recorded, along with information identifying the location in the program, and the selected presentation mode of "Speech Only" or "Speech & Text." In addition to itemizing the user choices, the data record for each user included the program running time, and totals for the four choice options of CONTINUE, Speech Only; CONTINUE, Speech & Text; REPEAT, Speech Only; and REPEAT Speech & Text.

As a second measurement instrument, each student completed a questionnaire designed to assess user response to speech displays and components of the program related to speech displays. The questionnaire included questions which addressed:

previous experience with hypermedia and with speech in hypermedia;

general reaction to the use of speech in the test program;

presentation preferences (speech vs. text) for verbal information/instruction in the test program;

overall reaction to the program "Field Kit Workshop."

Procedures

Development and testing of the program was conducted on an Apple Macintosh IIsi computer with high resolution 13-inch color monitor, 5 MB RAM and 40 MB internal hard drive. A small external amplifier and speaker were used for sound rather than the system's built-in speaker. The external amplifier allowed each user to easily set the program volume for his or her own comfort.

Thirteen individual sessions were conducted with the program over a period of four days. Three of these sessions, conducted on the first day of testing, were used to debug the program, and did not directly contribute data to this study. Based on these test runs of the program, some revisions were made to program delivery and navigation, and serious problems with the method of recording user activity were resolved. The ten sessions conducted after these revisions were made contributed the data for this study.

Upon arrival for testing, a participant was provided with a questionnaire and a manila envelope, and took his or her place at the computer. The researcher showed the participant the volume control, and, if necessary, provided a brief demonstration of using a mouse as input device to point, click and drag. The participant was then directed to begin. Introductory modules within the program itself provided information needed to use the program and to com-plete the questionnaire.

Because the program was still in a developmental stage, and not entirely free from bugs, the researcher remained in the vicinity during each session to troubleshoot any problems with the hardware or software. No direct observations of user behavior were made or recorded as a part of this study. It became obvious once the study was under way that direct observation of behavior would have provided additional data very useful as a component of formative evaluation; unfortunately, approval of this project by an oversight committee was based on a guarantee of

participant anonymity which could not be maintained if participant behavior was directly observed.

RESULTS

CHARACTERISTICS OF PARTICIPANTS

Half of the ten participants reported that they had never used a hypermedia program before. Of the five who had previous experience with hypermedia, four had used at least one program that presented information through the medium of speech.

On a scale from 1 to 5, 70% of the participants reported a level of experience with audio or video production equipment in general of either 4 or 5. A range of experience with video field production equipment specifically was more evenly distributed, with 40% reporting 1 or 2, 20% reporting 3, and 40% reporting 4 or 5 (see Figure 9).

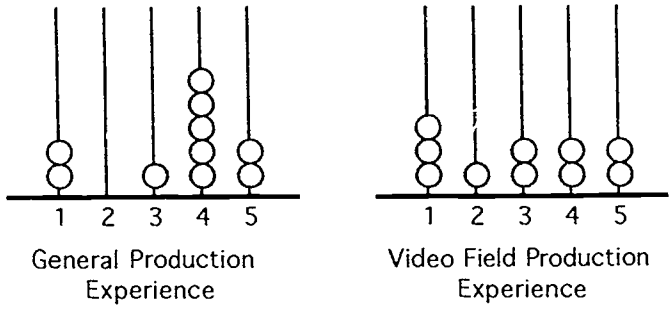

Figure 9. Reported Levels of Experience

Half of the participants had used the piece of equipment that was the subject of the program at least once.

SPEECH ONLY VS. SPEECH & TEXT

The preferred mode of presentation was Speech Only: seven of ten participants selected Speech Only more than 90% of the time. Only two of these participants reported having previously used hypermedia to receive information, instruction or training.

Three of the seven participants who demonstrated a preference for Speech Only presentation did vary somewhat the mode of presentation over the course of the program. One participant used Text & Speech for the first two nodes, and then switched to Speech Only for the entire remainder of the program. One used Speech Only throughout the program, and then switched to Text & Speech for the last two nodes. One student used Speech Only throughout the program, with one exception. In one node the user repeated a procedural instruction once as Speech Only, then switched to Speech & Text for a second repeat. After this second repeat, the user completed the requested task and returned to Speech Only mode to continue the program.

Among the three participants who demonstrated a preference for Speech & Text presentation, there was no variation from that mode. These three participants all reported having previously used hypermedia to receive information, instruction or training at least once; and all of these users had used the Sony BVU-150, the subject of the program, at least once.

The average level of agreement with the statement that speech "seemed **natural**, and was an effective way to receive instructions and information," was 4.0, on a scale from 1 to 5 where 1 =

"disagree" and 5 = "agree." 50% of the participants responded with the mode of 5, and 80% responded either 4 or 5. One participant responded 1, and one responded 2 (see Figure 10).

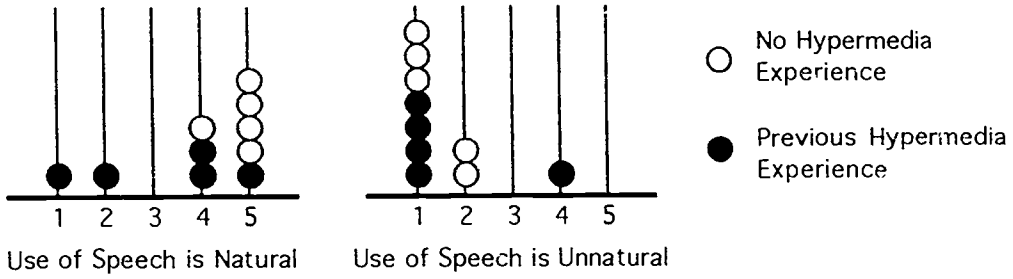

Figure 10. Use of Speech is Natural

In response to the question, "How much did the **use of speech** enhance your level of enjoyment of the program?" with 1 being "none," and 5 being "very much," the mean was 4.1, with 80% of the respondents giving ratings of either 4 or 5 (see Figure 11).

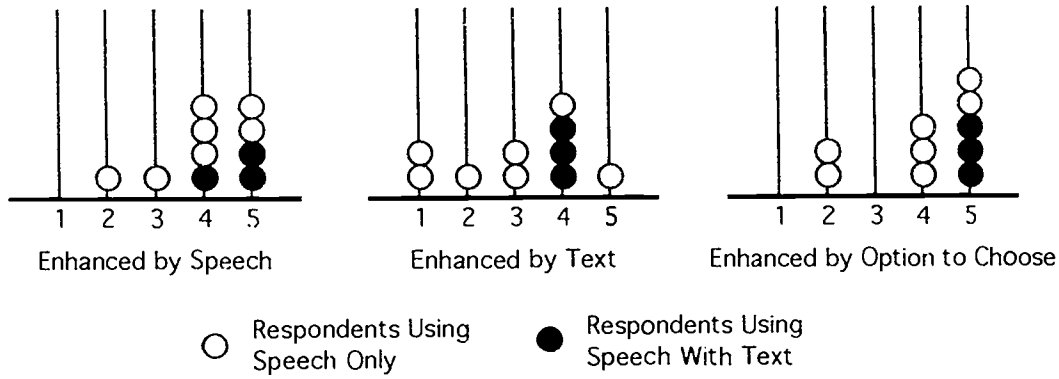

Figure 11. Responses to Use of Speech

80% of the respondents gave a rating of 4 when asked how easy it was to understand spoken instructions, with 1 being "very difficult" and 5 being "very easy." The mean was 4.0; 3 was the lowest rating received. The mean for ease of understanding written instructions was higher, at 4.4 (see Figure 12).

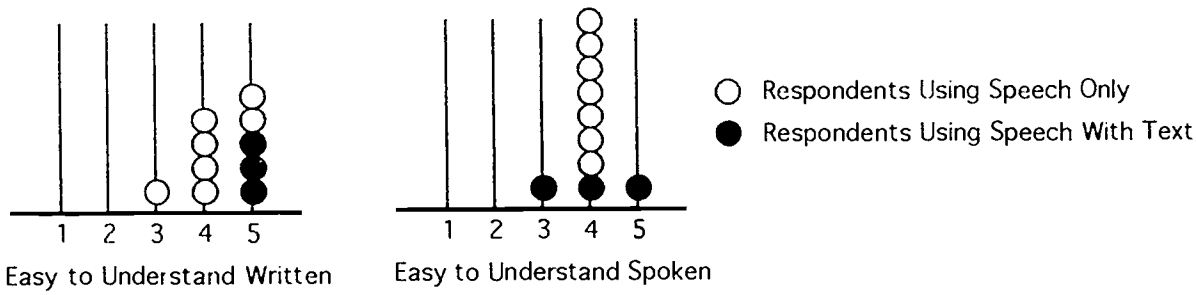

Figure 12 Understanding Speech and Text

There was greater agreement that the program would be improved if the "instructor" spoke more rapidly than there was that the program would be improved if the "instructor" spoke more slowly, although both suggestions received very low ratings: 1.6 was the mean for slower rate of speech, and 2.2 was the mean for faster rate of speech (where 1 = "disagree" and 5 = "agree"). The statement that the program would be improved if there were a variety of speakers throughout the program also received a low level of agreement, with a mean of 2.0.

Participants were presented five statements that described possible ways to use the Text Window within the program, and were asked to indicate any that described their own use. In keeping with the recorded data, 60% indicated that they "did not use the text window;" two participants (20%) indicated the statement that "Displaying the TEXT WINDOW helped me avoid having to use the REPEAT feature;" one indicated the statement that "With the TEXT WINDOW displayed, I sometimes missed details presented in visual images and animated sequences;" one indicated the statement, "Although I often displayed the TEXT WINDOW, I only referred to it occasionally:" and one indicated the statement, "Even with SPEECH, I depended mostly on the TEXT WINDOW for information."

In the course of the program, the user encountered a minimum of 47 prompts to continue or repeat (more if the user repeated). The mean number of repeats in Speech Only mode was 1.2; the mean number of repeats in Speech & Text mode was .3. The mean number of total repeats per participant was 1.5.

USING AND LEARNING

On a scale from 1 = "very difficult" to 5 = "very easy", the rating for overall ease of use had a mean of 4.6, with 60% of the responses being 5. Other use-related items on this scale included ease of operating controls, with a mean of 4.4; and ease of moving forward or backward through the program, with a mean of 4.3. (see Figure 13).

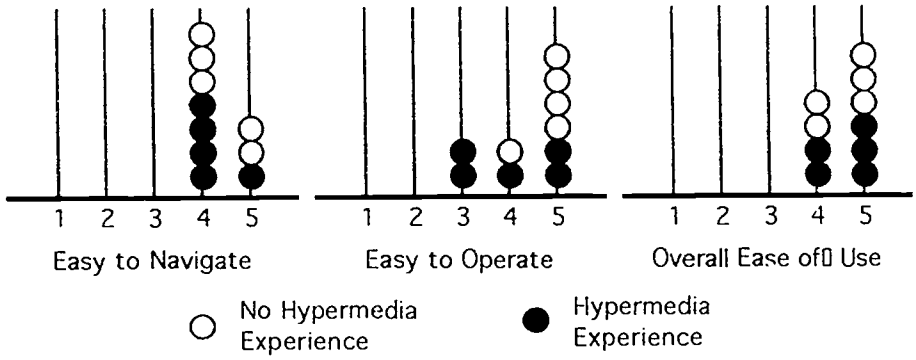

Figure 13. Ease of Use

When asked how easy it was to learn from the program, 50% of the participants assigned the highest rating of 5, with a mean of 4.4. All respondents reported that they had learned something new about the video tape recorder (VTR) in at least one of twelve listed content areas. The average number of content areas in which something was learned was 3.1. Among those who had previous experience with this particular VTR, the mean was 2.0; among those with no previous experience with the VTR, the mean was 4.2.

Asked "How confident are you that you have a **basic** understanding of how to operate the Sony BVU-150 video tape recorder," on a scale of 1 = "not confident" to 5 = "very confident," the mean for all responses was 4.3. Among participants who had used the VTR before, the mean was 4.6; among those who had not, the mean was 4.0.

When the participants were asked how much benefit they might receive from using the program a second time (on a scale from 1 = "none" to 5 = "very much"), the mean for all responses was 2.6. Against the same scale, when asked how much benefit would be received from having the program readily available for repeated use, the mean was higher, at 3.5 (see Figure 14).

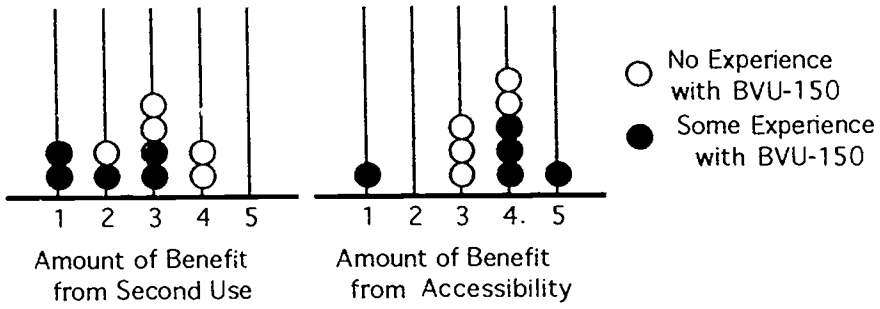

Figure 14. Expected Level of Benefit from Repeated Use

Participants were asked their preferred means of receiving a first introduction to a new piece of production equipment. In three separate items, 100% indicated a preference for using a hypermedia program over reading the equipment manufacturer's Operating Manual; 90% preferred using a hypermedia program over viewing a videotaped demonstration of the equipment; and 90% preferred using a hypermedia program over attending a small-group demonstration session (no hands-on) conducted by an experienced operator.

An overall level for enjoying the program was rated on a scale from 1 = "none" to 5 = "very much." 50% of the respondents gave the program the highest rating of 5; the mean was 4.3. Asked to rate, on the same scale, specific features that may have enhanced the level of enjoyment, the response mean for "realistic sound effects" was 4.3; for "use of speech" was 4.1; and for "use of music" was 3.2. The rating for the "option to choose" Speech Only or Speech & Text had a mean of 4.1; and for "use of text," the mean was 3.1. The rating for the "quality of the visuals" in enhancing the level of enjoyment had a mean of 4.3 (see Figure 15).

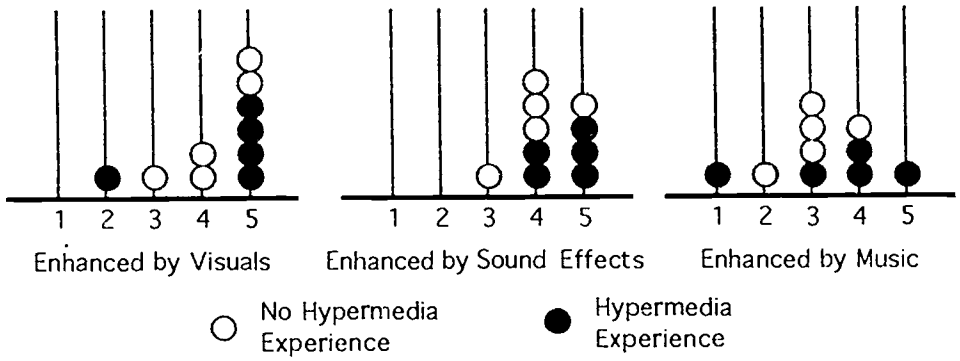

Figure 15. Other Enhancing Features

All respondents agreed with a statement that programs similar to the one tested should be developed for introducing students to the operation of other audio and video production equipment. On a scale with 1 = disagree and 5 = agree, all ratings were either 4 or 5; the mean was 4.3.

DISCUSSION

THE PARTICIPANTS

Given the small sample size, it was fortunate for this study that participants represented a range of experience with hypermedia and with video field production. The nearly even split of experienced and not experienced, across both categories, makes it possible to examine the data in ways not fully anticipated in the initial design.

It should be noted that participants reported a higher rate of previous exposure to speech in hypermedia than was expected, given that speech in hypermedia is not common. This high exposure is likely due to the fact that the sample was drawn from a population of students at a university that is active in developing and implementing hypermedia, and where there is a focus among developers on integrating sound into hypermedia programming.

SPEECH IN "FIELD KIT WORKSHOP"

The main purpose of this study as formative evaluation was to gather feedback to support the use of speech alone as a means of delivery for this particular program; a second aim was to gain insight into design factors that may have an impact on the effectiveness of speech delivery.

User Acceptance

The participants in this project did accept speech as a means of delivery. A strong majority chose the Speech Only mode of presentation, and even those who used the program with text support responded favorably to questionnaire items which addressed the use of speech.

The high rate of approval by participants suggests that a complete version of the program, "Field Kit Workshop," in which speech is the default and perhaps only mode of presentation for verbal information, could be designed to be effective, and would be accepted by the majority of those who would use the program. Nevertheless, enough participants took advantage of the option for text support to suggest that a text display option should be maintained.

When the data regarding use and acceptance of speech displays are viewed in terms of the users' previous exposure to hypermedia, an interesting trend is observed. As noted above, all of those who consistently selected Speech with Text as the mode of presentation reported having previous exposure to hypermedia; and the statement that speech seemed a natural way to receive information received it's lowest rates of agreement from two users who had previously used hypermedia.

As noted in the review of literature, computer-based instruction has traditionally delivered verbal information as text. While the data in this study is not conclusive, there is a suggestion that experienced hypermedia users have a positive bias toward the use of text, as a result of their past experience with computer-based delivery.

Speech and Understanding

It was beyond the scope of this study to provide a direct measure of the effectiveness of speech as a mode of delivery. Still, most users reported that speech was easy to understand; and the very low figures for repeats within the program support the notion that information was understood by all users, with or without text.

The low number of repeats, however, may have been the result of a low level of motivation to learn the material. Participants in the study would not necessarily be expected to ever use the piece of equipment that was the subject of the program, and so motivation to learn the material may have been low. The fact that only two of the fifteen repeats were repeats of entire nodes, while the remainder were repeats of only the procedural instructions, would seem to bear this

out. Some users may have been unclear about tutorial information and simply not bothered to repeat it, but the program was structured such that procedural instructions had to be understood before the user could continue.

Of the fifteen repeats that did occur, five were within one particular node within the program. The procedural instruction in this action node calls for the user to complete two actions in succession. This design is inconsistent with the rest of the program, in which each procedural instruction requires only one action.

Data that describe the number of repeats within this node are not good data because some users were told to repeat. But how these users repeated -- with Speech Only, or with Speech & Text -- is still useful data. When users repeated, did they choose a different presentation mode than they did for forward navigation through the program? If users who demonstrated a preference for Speech Only chose to REPEAT in Speech & Text mode, it would seem to indicate that these users thought the addition of text would improve the likelihood of understanding the instruction the second time. In fact, one user repeated the instruction one time as Speech Only, and then a second time as Text & Speech, before successfully completing the action. But for the most part, what was demonstrated was a strong tendency for users to use their preferred mode of presentation for REPEATS as well as for forward navigation.

While the repeat function was not heavily used, it did seem to serve the purpose of clarifying information for the user. Out of fifteen repeats, only twice did any user repeat the same chunk of speech twice. For all other instances, one repeat was sufficient to enable the user to proceed with the program.

Speech Characteristics

The low level of agreement with suggestions to increase or decrease the rate of speech seems to indicate that the decision to target 200 wpm as the average rate of speech for the program was a good one. And, while the designer had at one time considered using more than one voice through the course of the program, users did not feel that such an approach would add anything to the program.

THE PROGRAM

Speech was accepted as a medium within a program in which many other related and complementary components also received high approval ratings by users. The quality of the visuals and the use of realistic sound effects were also very well received. The use of music received a somewhat neutral response.

Overall, "Field Kit Workshop" received overwhelming approval as a training tool. After using FKW, most participants in the study indicated hypermedia as a preferred means for receiving initial equipment training, and all felt that programs similar to FKW should be developed for training students in the operation of other production equipment.

The only measure of the effectiveness of the program overall was the participants' own reporting. It came as no surprise that inexperienced participants reported learning more about the video tape recorder than experienced users did; it was somewhat of a surprise that all users reported learning something about the VTR -- even those who indicated a high level of experience with the Sony BVU-150.

SUMMARY

As outlined above, it was felt that a useful evaluation of speech in hypermedia could only be accomplished within a program that was well-designed overall. The high ratings this program received across all measures indicate that the project was successful in placing speech within an appropriate vehicle for examination.

This study found that speech will be accepted by users within a program that is well-designed overall, and in which the design takes into account the special strengths and weaknesses of speech as a medium for delivery.

It also found that users were generally satisfied with a speaking rate of approximately 200 words per minute. The high ratings for understandability of speech also suggest that a sampling rate of 11 kHz may be sufficient for recording speech, if care is taken in considering other recording factors, such as microphone selection.

RECOMMENDATIONS

THE FINAL DESIGN

The results of the evaluation supported the notion that speech could be used effectively to present information in this particular simulation. The final version of "Field Kit Workshop" will incorporate revisions in several areas to take full advantage of speech as a primary source for verbal information.

Because thirty percent of the users elected to receive text support for that narration, and eighty percent reported that the option to choose the mode of presentation enhanced their enjoyment of the program, the Speech & Text option will be maintained in the final design. But the way in which the option is offered will be revised.

In the prototype version of FKW, the user was required to make the decision of "Speech Only" or "Speech & Text" in conjunction with every navigation command to move forward or repeat. This was a design aimed at generating data for this study, and was not designed for the users' convenience. In the final version, the option to present text along with speech will be maintained, but the choice of mode will be made independently of navigation decisions. By removing the presentation mode options from the Repeat and Continue panels, the navigation devices -- in particular, the Repeat function -- can be more fully developed.

In the final version of the FKW, the user who is paused at an action point will be able to REPEAT either the procedural instruction alone, or can repeat back to the beginning of the node to receive the tutorial information as well as the procedural instruction. The final design for navigation and presentation panels is illustrated in Figure 16.

Paused at an Action Point, the user can repeat back to the beginning of the node, or just back to the procedural instruction. There is no option to continue at an Action Point.

Paused at a Navigation Point, the user can repeat back to the beginning of the recently completed node, or can continue to the next.

Figure 16. Redesigned Navigation & Presentation

A majority of the users indicated they would make additional use of "Field Kit Workshop" if it were readily available. The strictly linear and sequential navigation of the prototype reduces the usefulness of the program if it is to be used as a reference to specific information. To make the program more useful for repeat users, a menu will be added at the bottom of the control panel to allow the user to jump to certain topics.

FURTHER STUDY

There is clearly much that needs to be learned about the application of speech in hypermedia programming in general -- even considering only the use of speech as applied in "Field Kit Workshop," there are many questions that this small study did not treat.

Is text necessary at all in FKW? The decision was made to include text as a display option in the final version of "Field Kit Workshop," because almost one-third of the users selected the text option and most users appreciated having the choice. But further study, aimed at measuring the relative effectiveness of Speech Only vs. Speech with Text, may find that Speech Only presentation results in more effective learning under the conditions present in FKW.

In FKW, the most important information is in the active display area of the screen, and not in the text. Through images and sound, the student learns what the deck looks like, where certain controls are, and how the machine responds. The student who reads the text at the bottom of the screen may miss details of animated visual displays. Text seems to have an authority which people find hard to resist -- as one person who tried the program in an early stage of its development said, "With the text there, I just *have* to look at it."

A next step in examining speech presentation as it is applied in "Field Kit Workshop" might be to design an experiment to answer questions of relative effectiveness of speech with or without text. Do users respond more quickly to procedural instructions when text is not present? When the instructor gives a procedural instruction -- "Turn the Power Switch on," for example -- does the user who is not reading text respond more quickly and accurately? If not having to read the text means that the user has a head start scanning the screen for the power switch, then this user should be able to act more quickly.

It may also be that users can learn more detailed information without text display than with. FKW regularly uses animated sequences to illustrate certain procedures and characteristics of the deck, because animation is the most direct way to present the information. If the user is reading the text description that accompanies the animation, then that user may be missing the primary source of information -- the animated sequence. An experiment designed to test recall of animated sequences, comparing Speech Only and Speech with Text groups, may demonstrate that text can interfere with learning in these situations.

Also worth pursuing is the possibility that experienced hypermedia users are slower than first-time users when it comes to accepting speech as the sole source for verbal information. Incorporating speech as a regular component in the hypermedia mix could help make hypermedia accessible to a broader range of users -- but if the established base of users are slow to accept speech, and if developers are slow to implement it, then hypermedia may be unnecessarily slow in developing to its full potential as a powerful tool of learning.

APPLYING HYPERMEDIA AND SPEECH IN PRODUCTION INSTRUCTION

As a detailed simulation of one specific, technically sophisticated piece of equipment, "Field Kit Workshop," stands as an example of how a manufacturer might develop materials that can be used to provide training support for its products. For the educator thinking about developing hypermedia programming to complement classroom or lab activities, FKW also provide an example of the effective use of digitized speech to support the presentation of visual material.

In the field of video production, hypermedia programming has great potential for teaching basic concepts of the discipline; concepts such as shot composition, lighting techniques, and shot sequencing. Teaching these areas by any method requires extensive use of visual material -- often there are concepts of physics that need to be illustrated, and there are always examples of good and bad video to be shown. New hypermedia programs that are developed for teaching in the field of video production -- and other areas where the principle content of the instruction is visual -- should use speech to present verbal information. If your picture is worth a thousand words -- why clutter it up with a couple dozen more?

APPENDIX

EXCERPT OF PROGRAM SCRIPT FOR "FIELD KIT WORKSHOP"

program location	screen image	audio	name of sound
navigate			
CD 52: switchDisplay		"Next, you need to set the timecode information for this tape. Right now the tape counter is displaying control track information. "	setTimeCode
		"Flip the switch next to the reset button to the TimeCode (TC) position."	switchDisplay
action: switch TC			
	counter display shows: "00:00:00"	"Notice that the display now shows six digits: for hours, minutes and seconds."	TCDisplay
		"It doesn't show the individual frame numbers of the timecode. But they will be on the tape."	noFrames
navigate			
CD 53: CD id 27084	time code generator panel door opens	"Below the counter is a panel that controls the timecode generator."	TCControl
navigate			
CD 54: TCUbit		"Make sure the switch in the lower right of this panel is set to the TC, or TimeCode, position."	TCcode
action: switch TC			
	switch to TC	sfx: click	*click
navigate			
CD 55: setTCrun	animation: demonstration of counter in free-run mode	"If you put the RUN switch into Free-Run, the timecode will generate continuously, even when you are not recording."	FRun
		"We want the time code to advance only when recording -- what's called Record-Run. Set the RUN switch in the Record-Run position."	RRun
action: switch F-RUN			
	counter stops advancing	sfx: click	
navigate			
CD 56 setTCgen		"If you want to read timecode from a pre-recorded tape, the next switch needs to be in the Playback position."	tcPB
		"But we're recording, so we need to Generate timecode. Put the playback-or-generate switch in the GEN position."	tcGEN
action: switch GEN			
	switch to GEM	sfx: click	
navigate			

REFERENCES

Alessi, S. M., and Trollip, S.R. (1991). Computer-Based Instruction: Methods and Development (2 ed.). Englewood Cliffs, N.J.: Prentice Hall.

Carlson, P. A. (1990). The Rhetoric of Hypertext. Hypermedia, 2, 109-131.

Failo, T., and DeBloois, M. (1988). Designing a visual factors-based screen display interface: The new role of the graphic technologist. Educational Technology, (August).

Fleming, M. L., and Levie, W.H. (1978). Instructional message design: Principles from the behavioral sciences (3rd ed.). Englewood Cliffs, NJ: Educational Technology Publications, Inc.

Knuth, R. A., and Brush, T.A. (1990). Results of the hypertext '89 design survey. Hypermedia, 2, 91-107.

Locatis, C., Charuhas, J., and Banvard, R. (1990). Hypervideo. Educational Technology Research & Development, 38(2), 41-49.

Marics, M. A., and Williges, B.H. (1988). The intelligibility of synthesized speech in data inquiry systems. Human Factors, 30(December), 719-732.

Zilber, J. (1992). The Agenda Gap. MacUser, 8(3), 25-26.